



Investigadores de la Politècnica de València desarrollan un nuevo método para la detección automática de plagios de texto

- El método, basado en modelos de traducción estadística, permite detectar plagio traducido
- Estos modelos “aprenden” cuáles son las potenciales traducciones de un texto de un idioma a otro a partir de grandes colecciones de documentos.

Expertos del Laboratorio de Ingeniería en Lenguaje Natural (Lab NLE), integrado en el Grupo de Ingeniería del Lenguaje Natural y Reconocimiento de Formas (ELiRF) de la Universitat Politècnica de València, han desarrollado un nuevo método para la detección automática de textos plagiados y, en concreto, de plagio traducido. Este trabajo ha sido publicado recientemente en “Languages Resources and Evaluation Journal” y se enmarca dentro del proyecto Text-Enterprise 2.0, financiado por el Ministerio de Ciencia e Innovación.

“En nuestros días, el problema del plagio, en particular el de texto, se ha incrementado debido al fácil acceso a grandes fuentes de información a través de medios electrónicos. Desafortunadamente, su detección es prácticamente imposible de forma manual. Por ello, es importante desarrollar mecanismos automatizados que permitan realizar la tarea de detección de plagio y así combatir la creciente tentación de plagiar desde la Web”, apunta Paolo Rosso, investigador del Lab NLE de la Politècnica de València.

Junto a Alberto Barrón-Cedeño, estudiante de doctorado de la UPV, Paolo Rosso trabaja actualmente en el desarrollo de tecnología y aplicaciones para la detección automática de plagio y concretamente de plagio traducido, por ejemplo cuando la fuente está escrita en inglés y el plagio está en castellano o valenciano.

“El plagio traducido se hace más frecuente cuando la información que se busca no está disponible en la Web en el idioma materno, por ejemplo el valenciano. Si la detección de plagio en una misma lengua es de por sí complicada, ya que una persona puede modificar el texto original, cuando esta modificación implica un cambio de lengua la dificultad es aún mayor. Pocos son los métodos que han sido desarrollados para abordar este tipo de plagio”, explica el profesor Rosso.

Los investigadores de la UPV han desarrollado uno de estos métodos, que se basa concretamente en modelos de traducción estadística. Estos modelos “aprenden” cuáles son las potenciales traducciones de un texto de un idioma a otro a partir de grandes colecciones de documentos.

“Con lo aprendido es posible calcular la similitud entre textos escritos en distintos idiomas y, si esta similitud es muy alta, se puede sospechar de la existencia de un caso de plagio traducido”, apunta Alberto Barrón-Cedeño.

A diferencia de otros modelos existentes, el método desarrollado desde el Laboratorio del NLE de la Politècnica de València tiene la ventaja de ser robusto cuando se enfrenta a lenguas poco relacionadas sintácticamente (por ejemplo, castellano y euskera).

“El objetivo es proporcionar la evidencia necesaria para que un experto tenga las mejores condiciones posibles para tomar una decisión final respecto a si ha habido o no plagio”, añade Paolo Rosso.



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Nota de prensa

Competición Internacional

En el marco de esta línea de investigación, el Lab NLE de la Universitat Politècnica de València ha organizado, conjuntamente con la Bauhaus Universitat Weimar (Alemania), y por tercer año consecutivo, una competición internacional sobre detección automática de plagio (<http://pan.webis.de/>).

Patrocinada por Yahoo! Research, esta competición tendrá lugar en Amsterdam el próximo mes de septiembre, dentro del foro europeo de evaluación del CLEF (<http://clef2011.org/index.php?page=papers/labs.html>),

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Área de Comunicación

Edificio Nexus (6G), Camino de Vera, s/n - 46022 VALENCIA



Investigadors de la Politècnica de València desenvolupen un nou mètode per a la detecció automàtica de plagis de text

- El mètode, basat en models de traducció estadística, permet detectar plagis traduït
- Aquests models “aprenen” quines són les potencials traduccions d'un text d'un idioma a un altre a partir de grans col·leccions de documents.

Experts del Laboratori d'Enginyeria en Llenguatge Natural (Lab NLE), integrat en el Grup d'Enginyeria del Llenguatge Natural i Reconeixement de Formes (ELiRF) de la Universitat Politècnica de València, han desenvolupat un nou mètode per a la detecció automàtica de textos plagiats i, en concret, de plagis traduït. Aquest treball ha sigut publicat recentment en *Languages Resources and Evaluation Journal*, i s'emmarca dins del projecte Text-Enterprise 2.0, finançat pel Ministeri de Ciència i Innovació.

“Actualment, el problema del plagi, en particular el de text, s'ha incrementat a causa de l'accés tan fàcil a grans fonts d'informació a través de mitjans electrònics. Desafortunadament, la detecció d'aquest plagi és pràcticament impossible de forma manual. Per això, és important desenvolupar mecanismes automatitzats que permeten dur a terme la tasca de detecció de plagi i així combatre la temptació creixent de plagiar des de la web”, apunta Paolo Rosso, investigador del Lab NLE de la Politècnica de València.

Juntament amb Alberto Barrón-Cedeño, estudiant de doctorat de la UPV, Paolo Rosso treballa actualment en el desenvolupament de tecnologia i aplicacions per a la detecció automàtica de plagis i concretament de plagis traduït, per exemple, quan la font està escrita en anglès i el plagi està en castellà o valencià.

“El plagi traduït es fa més freqüent quan la informació que se cerca no està disponible en la web en l'idioma matern, per exemple el valencià. Si la detecció de plagis en una mateixa llengua és per si mateix complicada, ja que una persona pot modificar el text original, quan aquesta modificació implica un canvi de llengua la dificultat és encara més gran. Pocs són els mètodes que han sigut desenvolupats per a abordar aquest tipus de plagis”, explica el professor Rosso.

Els investigadors de la UPV han desenvolupat un d'aquests mètodes, que es basa concretament en models de traducció estadística. Aquests models *aprenen* quines són les potencials traduccions d'un text d'un idioma a un altre a partir de grans col·leccions de documents.

“Amb el que s'ha après és possible calcular la similitud entre textos escrits en diferents idiomes i, si aquesta similitud és molt alta, es pot sospitar de l'existència d'un cas de plagi traduït”, apunta Alberto Barrón-Cedeño.

A diferència d'altres models, el mètode desenvolupat des del Laboratori de l'NLE de la Politècnica de València té l'avantatge de ser robust quan s'enfronta a llengües poc relacionades sintàcticament (per exemple, castellà i basc).

“L'objectiu és proporcionar l'evidència necessària perquè un expert tinga les millors condicions possibles per a prendre una decisió final respecte a si hi ha hagut plagis o no”, afegí Paolo Rosso.

Competició internacional



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Nota de prensa

En el marc d'aquesta línia d'investigació, el Lab NLE de la Universitat Politècnica de València ha organitzat, conjuntament amb la Bauhaus-Universität Weimar (Alemanya), i per tercer any consecutiu, una competició internacional sobre detecció automàtica de plagi (<http://pan.webis.de/>).

Patrocinada per Yahoo! Research, aquesta competició tindrà lloc a Amsterdam el pròxim mes de setembre, dins del fòrum europeu d'avaluació del CLEF (<http://clef2011.org/index.php?page=pages/labs.html>).

Datos de contacto:

Luis Zurano Conches
Unidad de Comunicación Científica e Innovación
Universitat Politècnica de València
Mòvil: 647 422 347

Anexos:

Datos de contacto:

Luis Zurano Conches
Unidad de Comunicación Científica e Innovación
Universitat Politècnica de València
Mòvil: 647 422 347

Anexos:

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Àrea de Comunicació

Edificio Nexus (6G), Camino de Vera, s/n - 46022 VALENCIA